EX. G

Red-Shifted Excitation Mutants of the Green Fluorescent Protein

Simon Delagrave[1], Rachael E. Hawtin, Christopher M. Silva, Mary M. Yang[2] and Douglas C. Youvan*

Palo Alto Institute of Molecular Medicine, 2462 Wyandotte St., Mountain View, CA 94043, USA. [1]Present address: Howard Hughes Medical Institute, Dept. of Biology, Brandeis University, Waltham, MA 02254. [2]KAIROS Scientific Inc., 176 Covington Court, San Jose, CA 95136. *Corresponding author (e-mail: YOUVAN@AOL.COM).

Using optimized combinatorial mutagenesis techniques and Digital Imaging Spectroscopy (DIS), we have isolated mutants of the cloned *Aequorea victoria* green fluorescent protein (GFP) that show red-shifted excitation spectra similar to that of *Renilla reniformis* GFP. Selective excitation of wild-type versus Red-Shifted GFP (RSGFP) enables spectral separation of these proteins. Six contiguous codons spanning the tyrosine chromophore region were randomized and sequence analysis of the mutants revealed a tyrosine-glycine consensus. These mutants will enable the simultaneous analysis of two promoters or proteins per cell or organism. In consideration of the multitude of applications which are developing for GFP alone, we envisage that spectrally shifted fluorescent proteins will be of value to a diversity of research programs, including developmental and cell biology, drug-screening, and diagnostic assays.

Received 2 December 1994; accepted 22 December 1994.

The green fluorescent protein is the ultimate source of fluorescent light emission in the jellyfish *Aequorea victoria*. GFP has been cloned[1], leading to its widespread use as a marker for gene expression and as a protein tag, both in cell culture and in multicellular organisms[2]. GFP absorbs blue light at 395 nm; its excitation in jellyfish has been proposed to occur by radiative energy transfer from the photoprotein aequorin[3]. Wild-type GFP has a fluorescence emission maximum of 510 nm. The protein is easily isolated, very stable, and forms crystals which diffract to 2.2 Å resolution[4].

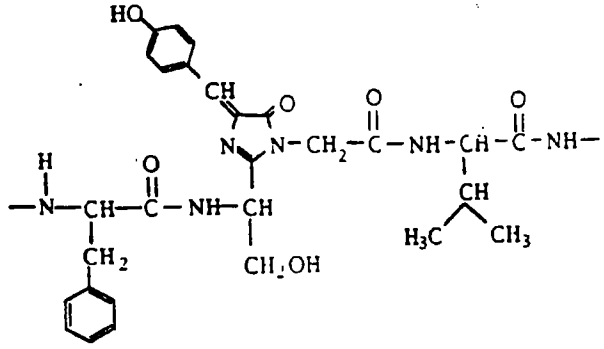
Another naturally occurring green fluorescent protein has been found in the sea pansy *Renilla reniformis*. Although the *Renilla* GFP gene has not yet been characterized, the protein has been studied in some detail. The emission spectrum of this protein is almost identical to that of *Aequorea* GFP. However, the two GFPs are easily distinguished because their excitation maxima are red-shifted by over 100 nm, from 393 nm in *Aequorea* GFP to 498 nm in *Renilla* GFP. There is strong evidence that *Renilla* and *Aequorea* GFP carry the same chromophore[5].

A putative GFP chromophore structure (Fig. 1) was described in 1979, and has since been modified[6,7]. This structure is unique amongst chromophores, comprising a cyclic tripeptide sequence, serine-dehydrotyrosine-glycine, which is covalently linked through the protein's peptide backbone. The mechanism of chromophore formation is unknown, but oxidation of tyrosine to dehydrotyrosine is probably essential.

The methods used for the construction of spectrally shifted GFP mutants have been successfully employed in the past[8,9] to produce a variety of spectrally diverse bacteriochlorophyll-binding proteins using optimized combinatorial mutagenesis and Digital Imaging Spectroscopy (DIS). DIS enables simultaneous screening of thousands of colonies directly on petri dishes by acquiring spatially resolved spectral information[10-14]. Images of petri dishes, illuminated at different wavelengths, are captured by a charge-coupled device (CCD) camera and further processed by software establishing radiometric calibration. Using optimized combinatorial mutagenesis and DIS, we describe the construction, isolation and fluorescent properties of several red-shifted GFP mutants.

Results

Combinatorial library screening. The region of GFP targeted for mutagenesis is the 6 amino acid sequence between phenylalanine 64 and glutamine 69 (FSYGVQ) which includes the chromophore itself. A mutagenic oligonucleotide (see Experimental Protocol) was designed to favor the incorporation of an aromatic amino acid at position 66 and to fully randomize the other five codons. The mutagenic portion of the oligonucleotide includes the sequence: NNK NNK TDK NNK NNK, where mixed nucleotides are denoted by N (all four nucleotides), K (G and T), and D (A, G and T). The sequence of the oligonucleotide employed for mutagenesis was obtained using the CyberDope computer program[15,16]. This software establishes a table of all possible codon nucleotide mixtures (or dopes) and calculates the frequency at which the corresponding



(48)CTTGKLPVPWPTLVTTFSYGVQC(70)

FIGURE 1. Structure of the GFP chromophore and nearby sequence. The tyrosine residue shown in the sequence is actually the dehydrotyrosine component of the chromophore. The tryptophan residue is thought to interact with the chromophore by energy transfer.

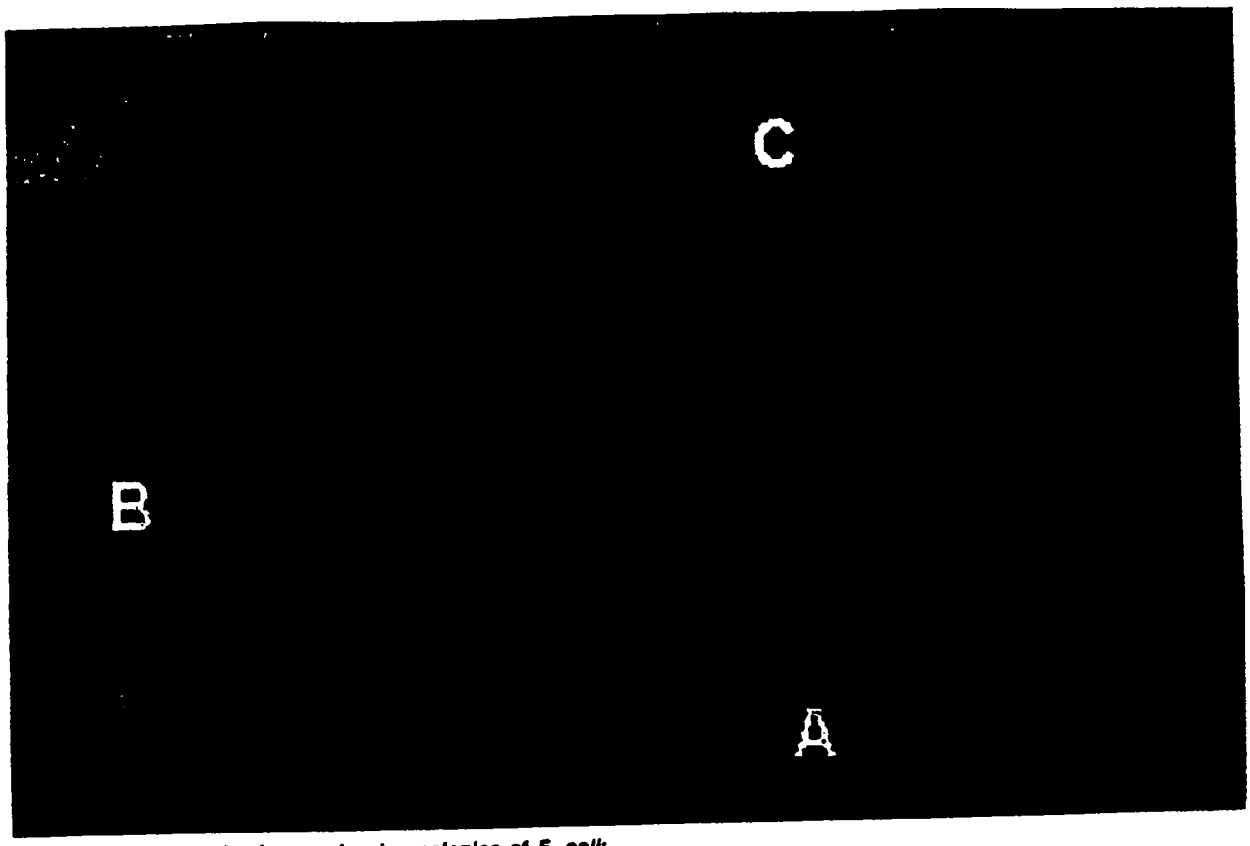


FIGURE 2. Pseudocolor image showing colonies of *E. coli*: (A) not expressing GFP, (B) expressing wild-type GFP, and (C) expressing a red-shifted excitation mutant of GFP.

amino acids will be encoded. The program then compares the table with a set of amino acids specified by the experimenter and outputs the codon nucleotide mixture which approximates that subset most closely. Therefore, doping schemes can be formulated which encode specific subsets of amino acids and exclude others. This maximizes the efficiency of semi-random mutagenesis, as selected dopes can be introduced into a random sequence to increase the probability of obtaining functional mutants[14]. For example, the probability of encoding tyrosine in a random codon mixture (e.g., NNK) is 0.03 compared with 0.17 from the optimized codon (TDK), in which TAT encodes Tyr.

The mutagenic oligonucleotide was used with a second primer (see Experimental Protocol) in an inverse PCR mutagenesis scheme[10]. Plasmid pTU58 is a pET3a derivative expressing GFP under the control of a T7 promoter[2]. The linearized plasmid was amplified in a PCR using both primers. The PCR products, which all incorporate the sequence of the mutagenic oligonucleotide, are circularized by ligation and transformed into *E. coli*. The resulting library of approximately 3×10^4 mutant GFP genes was expressed in BL21 (DE3). Thousands of colonies on petri dishes were screened by fluorescence using DIS. The spectrally shifted mutants were initially identified by their green fluorescence observed when excited with 490 nm light, which disappears when excited at 410 nm. In contrast, wild-type GFP fluorescence is much brighter with 410 nm illumination. DIS revealed that approximately one in 10^4 colonies expressed a functional fluorescent protein.

TABLE 1. Red-shifted GFP sequences resulting from combinatorial mutagenesis near the chromophore, including tyrosine 66. Wild-type *Aequorea* GFP and *Renilla* GFP chromopeptide sequences are also shown.

Strain	Sequence Position 64	65	66	67	68	69
Aequorea	F	S	Y	G	V	Q
Renilla	F	S	Y	G	D	R
RSGFP1	G	S	Y	G	L	L
RSGFP2	L	L	Y	G	A	Q
RSGFP3	G	C	Y	G	M	N
RSGFP4	M	G	Y	G	V	L
RSGFP6	V	A	Y	G	M	L
RSGFP7	L	C	Y	G	T	V

Several red-shifted GFP (RSGFP) clones were picked and sequenced (Table 1). Tyr^{66} and gly^{67} appear to be conserved while the other four positions are less stringent; ser^{65} is not necessary for the observed phenotype. Substitutions involving charged residues or aromatic residues were not observed in this limited data set, in contrast with the wild-type *Renilla* GFP hexapeptide sequence[15]. To verify that the combinatorial mutations were the cause of the red-shift, RSGFP4 was reconstructed by subcloning a 220 bp sequenced fragment of the mutant gene into a wild-type GFP expression vector (pTU58K). The resulting construct has the same spectral properties as the original RSGFP4 isolate, confirming that the mutations shown in Table 1 cause the red-shifted phenotype. This sequence information will be amenable to further manipulation by Exponential Ensemble Mutagenesis (EEM) and Recursive Ensemble Mutagenesis (REM) strategies[9,11], potentially to produce a 'rainbow' of multispectral fluorescent proteins. Specifically, we expect that by constructing new combinatorial libraries optimized by REM or EEM, the frequency of functional mutants will be high enough to allow the isolation of rare clones with significant emission shifts.

Spectral characterization of mutants. As a practical demonstration of the spectral separability of GFP and RSGFP, a

RGB (pseudocolor) image (Fig. 2) was generated by DIS. A petri dish was streaked with three strains of *E. coli* expressing (A) a non-functional GFP mutant, (B) wild-type *Aequorea* GFP and (C) RSGFP7. In this pseudocolored image, fluorescence images have been loaded into the red and green channels and an absorption image has been loaded into the blue channel. Both fluorescence images were captured using a 510 nm bandpass filter with excitation either at 410 nm ('green' channel) or 470 nm ('red' channel). These three monochrome images were combined into a 24-bit RGB image as described in the Experimental Protocol.

The absorption spectrum of partially purified RSGFP4 protein (data not shown) as well as its fluorescence excitation and emission spectra (Fig. 3A), are very similar to that of *Renilla* GFP. The fluorescence excitation and emission maxima of RSGFP4 are 490 nm and 505 nm, respectively, as seen in Figure 3A. While the emission of RSGFP4 is nearly identical to that of wild-type GFP (Fig. 3B), the excitation spectra are very different.

Discussion

We have isolated RSGFPs which are easily distinguished from wild-type GFP because their excitation maxima are red-shifted by about 100 nm, from 390 nm in wild-type *Aequorea* GFP to 490 nm in RSGFP. The spectral properties of these red-shifted mutants are very similar to those of *Renilla* GFP, corroborating the idea that *Aequorea* and *Renilla* GFP have identical chromophores. One might hypothesize that the microenvironment of the RSGFP chromophore may have become less acidic, making the deprotonated form of the chromophore the predominant absorbing species. This would be in line with chromophore titration experiments on wild-type GFP showing a clear transition from maximal absorption at 390 nm to 450 nm as pH is increased[3]. A biochemical and biophysical investigation of these mutants should clarify the exact molecular basis of the observed spectral changes.

It has been shown that recombinant GFP with fluorescence excitation and emission spectra identical to that of the native protein can be expressed in a wide variety of organisms, including: bacterial cells, *Caenorhabditis elegans*[2] and COS cells[11]. Expression of the protein from heterologous promoters enables analysis of the timing of gene expression without the requirement for addition of exogenous substrates. This is exemplified by the expression of GFP under the control of a heterologous promoter in neurons of the worm *C. elegans*. In addition to facilitating the analysis of gene expression, the production of fusion proteins comprising GFP and a protein of interest may enable tracking of the protein *in vivo* in transparent organisms such as the zebra fish or *C. elegans*.

Co-expression of GFP and RSGFP will enable the analysis of two proteins or promoters per cell or organism. The fluorescence imagery shown in this work demonstrates the feasibility of spectrally separating these two different tags. This makes it possible to analyze gene expression cascades, monitor the effect of drugs on the expression of a range of proteins within the cell, or to simultaneously track more than one protein within a living organism.

It has been reported that GFP fusion proteins can be expressed in *Drosophila melanoguster* and that fluorescence is detectable in thin sections following formaldehyde fixation[2]. Our macroscopic imagery of colonies expressing GFP and RSGFP on petri dishes is spectrally analogous to the problem of resolving these two tags in microscopic images. RSGFP and other spectrally separable derivatives of GFP should therefore be of value to confocal microscopy, flow cytometry, and fluorescence activated cell sorting using modular flow, dual excitation techniques. Further, the RSGFP excitation spectrum is close to optimal for Argon ion laser excitation at 489 nm. In the future, GFP derivatives may be engineered to respond to intracellular ion concentrations and other small molecules.

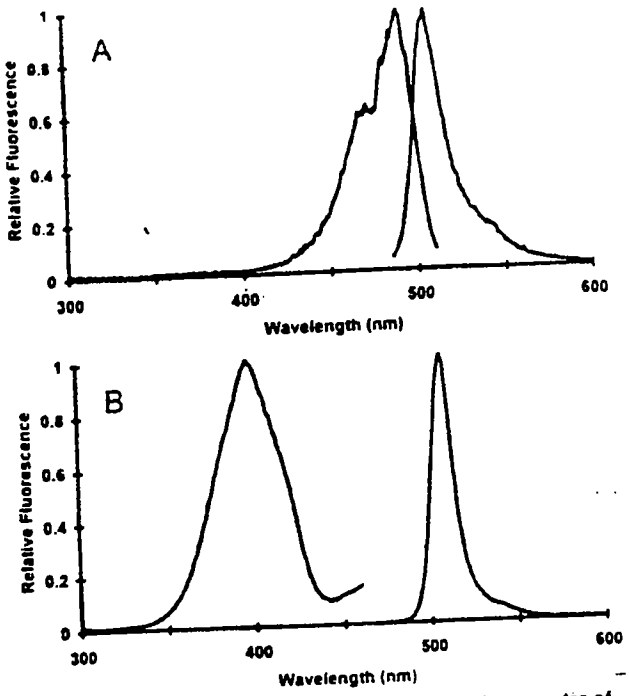


FIGURE 3. (A) Fluorescence excitation and emission spectra of mutant RSGFP4. (B) Fluorescence excitation and emission spectra of wild-type *Aequorea* GFP. Note that the excitation spectrum of RSGFP is red-shifted almost 100 nm relative to wild-type GFP.

Experimental Protocol

DNA manipulations and mutagenesis. Oligonucleotides were obtained from Operon Technologies (Alameda, CA). Restriction enzymes and T4 DNA ligase were from Gibco-BRL (Grand Island, NY) and New England Biolabs (Beverly, MA). DNA sequencing was carried out with a Sequenase 2.0 kit from USB (Cleveland, OH) on double-stranded templates purified with a Qiagen (Chatsworth, CA) spin-prep kit. A complex combinatorial library was constructed in *E. coli* BL21 (DE3) which expresses T7 RNA polymerase. A pET3a derivative (pTU58, provided by M. Chalfie) expresses high levels of GFP in this strain. Plasmid (pTU58) DNA was linearized by restriction with Nco I and amplified by an Inverse PCR method[10] using Taq polymerase, the PCR Extender kit (Stratagene (La Jolla, CA)) and two primers. PCR was performed on a Coy Tempcycler (Manufacturer, City, State) with the following parameters: 30 cycles of 95°C for 1 min, 45°C for 30 s, 75°C for 6 min and a 5 min final extension at 75°C. Mutations due to PCR are expected at a frequency of about 1 in 700 bp and have been observed in sample sequences of the clones. For example, RSGFP2 bears a Phe[71]->Leu point mutation. The mutagenic oligonucleotide had the following sequence (Nco I site underlined): 5′ AAACTACCTGTTCCATGGCCAACACTTGTCACTACTN-NKNNKTDKNNKNNKNNKTGCTTTTCAAGGTACCCAGATC. The other PCR primer had the following antisense sequence: 5′ GACAAGTGTTGGCCATGGAACAGGTAGTTTTCCAGTAGTGC. The reaction product comprises the entire plasmid containing the mutagenized sequence encoded by the primers, and Nco I sites at both ends of all molecules. PCR products were gel-purified, restricted with Nco I and self-ligated. The ligation product was electroporated into XL1-Blue (Stratagene) competent cells. An aliquot of the pool of transformants was plated on LB^{Amp} plates to determine the number of transformants (3×10^5 mutants), while the remainder was grown over-night in LB^{Amp}. DNA was purified from the pool of transformants and approximately 0.1 μg was used to transform BL21 (DE3). The resulting library was plated and then screened by DIS.

Reconstruction of RSGFP4. A translationally silent Kpn I restriction site was engineered in pTU58 by inverse PCR mutagenesis. The resulting plasmid (pTU58K) was shown by DIS to express wild-type GFP. To reconstruct RSGFP4, a 220 bp Nhe I-Kpn I restriction fragment spanning the mutagenized region of the RSGFP4 sequence was gel purified. pTU58K was also digested with Nhe I and Kpn I and the large restriction fragment gel purified. The two fragments were ligated, thus replacing the wild-type chromophore sequence with the mutant sequence of RSGFP4.

The ligation product was used to transform BL21 (DE3). Transformants were analyzed by DIS, as described below.

Digital Imaging Spectroscopy. Plates (LB^{amp}) with colony densities ranging from 100 to 900 colonies/plate were grown over-night at 37°C. We have observed that higher levels of fluorescence require aging the plates at room temperature for an additional 24 hours. Digital Imaging Spectroscopy utilized the KAIROS Scientific Inc. *ColonyImager* (San Jose, CA), as previously described in applications on other genetically altered pigment-protein systems. Such methodology has been recently reviewed[13,14]. Using the *ColonyImager*, three 16-bit monochrome images were acquired for Figure 2: blue channel = 0.2 second exposure (absorbance image) at 650 nm, grayscale values of 876 to 2365; green channel = 30 second exposure with 410 nm excitation (bandpass) and 510 nm fluorescence emission filter (bandpass), grayscale values 57 to 305; red channel = 30 second exposure with 470 nm excitation (bandpass) and 510 nm fluorescence emission filter (bandpass), grayscale values 43 to 577. All bandpass filters are of the Fabry-Perot type with 10 nm width-at-half-height. A dark frame was subtracted from each image to correct for dark counts. To make the RGB (pseudocolor) image, the three images were thresholded to their lowest grayscale values, contrast enhanced on a 0 to 255 range (8 bit) and combined to form the 24 bit color image. CyberDIS software (KAIROS) was used to optimize the selection of a palette of colors.

Fluorimetry. Protein samples were obtained by French pressing 250 mL of a 48 hour culture (grown at 37°C in LB containing ampicillin 50 μg/mL) in lysis buffer[2]. Cell debris was pelleted by centrifugation and the supernatant brought to 35% saturated ammonium sulfate (AS). This solution was centrifuged and the supernatant (containing GFP) brought to 70% AS. GFP was pelleted from the resulting solution by centrifugation and solubilized in 25% AS. The protein was then applied to a G-50 sephadex column. The peak fraction was used for fluorescence spectroscopy. Excitation and emission spectra (Fig. 3) were taken on a Photon Technology Inc. fluorometer. Wild-type GFP was excited at 395 nm and emission was measured at 510 nm. Mutant RSGFP4 was excited at 470 nm and emission monitored at 520 nm.

Acknowledgments

We wish to thank David Fork, of the Carnegie Institute of Washington (Stanford, CA) for fluorescence spectroscopy. This work was supported by NIH GM42645, DOE DE-FG02-90ER20019, and by DOE DE-FG05-ER79031. S.D was partially supported by the Fonds FCAR. We also wish to thank Robert Baltz for helpful comments and discussions.

Note Added in Proof

Green fluorescent protein mutants have been reported recently showing significant excitation and emission shifts (Heim, R., Prasher, D. C. and Tsien, R. Y. 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. USA. 91:12501-12504).

References

1. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. and Cormier, M. J. 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene. 111:229-233.
2. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. and Prasher, D. C. 1994. Green fluorescent protein as a marker for gene expression. Science 263:802-805.
3. Ward, W. W. 1979. Energy transfer processes in bioluminescence. Photochem. Photobiol. Rev. 4:1-57.
4. Perozzo, M. A., Ward, K. B., Thompson, R. B. and Ward, W. W. 1988. X-ray diffraction and time-resolved fluorescence analyses of *Aequorea* green-fluorescent protein crystals. J. Biol. Chem. 263:7713-7716.
5. Ward, W. W., Cody, C. W. and Hart, R. C. 1980. Spectrophotometric identity of the energy transfer chromophores in *Renilla* and *Aequorea* green fluorescent proteins. Photochem. Photobiol. Rev. 31:611-615.
6. Shimomura, O. 1979. Structure of the chromophore of *Aequorea* green fluorescent protein. FEBS Lett. 104:220-222.
7. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G. and Ward, W. W. 1993. Chemical structure of the hexapeptide chromophore of the *Aequorea* green fluorescent protein. Biochemistry 32:1212-1218.
8. Goldman, E. R. and Youvan, D. C. 1992. An algorithmically optimized combinatorial library screened by digital imaging spectroscopy. Bio/Technology 10:1557-1561.
9. Delagrave, S. and Youvan, D. C. 1993. Searching sequence space to engineer proteins: Exponential ensemble mutagenesis. Bio/Technology 11:1548-1552.
10. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. and Galas, D. J. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Research 17:6546-6551.
11. Delagrave, S., Goldman, E. R. and Youvan, D. C. 1993. Recursive ensemble mutagenesis. Protein Engineering 6:327-331.
12. Inouye, S. and Tsuji, F. I. 1994. Expression of the gene and fluorescence characteristics of the recombinant protein. FEBS Lett. 341:277-280.
13. Youvan, D. C., Goldman, E. R., Delagrave, S. and Yang, M. M. 1995. Digital imaging spectroscopy for massively parallel screening of mutants. Methods in Enzymology 246:732-748.
14. Youvan, D. C. 1994. Imaging sequence space. Nature 369:79-80.
15. San Pietro, R. M., Prendergast, F. G. and Ward, W. W. 1993. Sequence of the chromogenic hexapeptide of *Renilla* green-fluorescent protein. Photochem. Photobiol. 57:63S.